A black background with white text

Description automatically generated with low confidence￼

MSc Data Science Project

7PAM2002-0509-2024

Department of Physics, Astronomy and Mathematics

**Data Science FINAL PROJECT REPORT**

**Project Title:**

“Enhancing Brain Tumor Classification with Machine Learning on T1-Weighted MRI: Comparing Custom and Transfer Learning Models "

**Student Name and SRN:**

MD fahim and Student ID:23104402

Supervisor: Dr. Vid Irsic ([v.irsic@herts.ac.uk](https://mailto:v.irsic@herts.ac.uk/))

Date Submitted: 28th September 2025

Word Count: 5019

Github Repository:

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science **in Data Science** at the University of Hertfordshire.

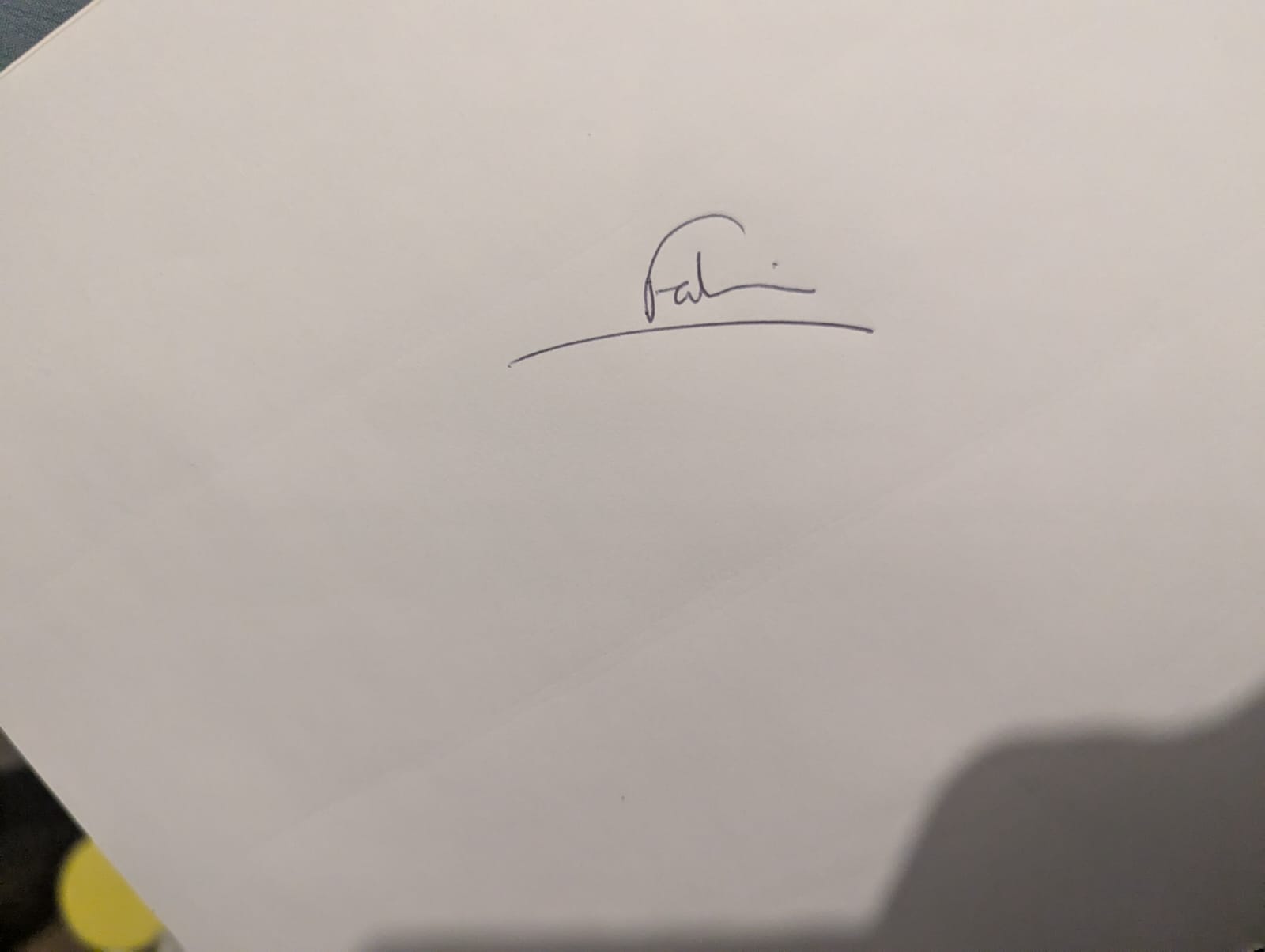
I have read the detailed guidance to students on academic integrity, misconduct and plagiarism information at [Assessment Offences and Academic Misconduct](https://ask.herts.ac.uk/assessment-offences-and-academic-misconduct) and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6)

I did not use human participants in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student Name printed: Md fahim



Student Name signature

Student SRN number:23104402

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

[**1. Abstract 4**](#_s4p4qy6acn0)

[**2. Introduction 4**](#_lpdegi4ai1kg)

[**2.1. Background and Clinical Context of Brain Tumor Diagnosis 5**](#_bai5ew5eq88s)

[**2.2. Challenges in Manual MRI Interpretation 5**](#_yabj795ghcu5)

[**2.3. Project Aims and Objectives 5**](#_kjkq80wq7ro4)

[**3. Literature Review 6**](#_bwi1na21nqkc)

[**3.1. Evolution from Traditional Machine Learning to Deep Learning 6**](#_fdnzrfueiwju)

[**3.2. Foundational Role of Convolutional Neural Networks 6**](#_azjqsygczjxk)

[**3.3. Architectural Review: EfficientNetB0 and ResNet50 7**](#_96egjswnpvuw)

[4. Materials and Methods 7](#_dxdzji966f4e)

[**4.1. Dataset Overview 7**](#_nkm5ox8izd1f)

[**4.2. Data Preprocessing 8**](#_u2ud7skqres9)

[**4.3. Data Ethics 9**](#_4pdxq7fellbe)

[**5. Methodology 9**](#_v07hwtxivdfs)

[**5.1. Data Preprocessing Pipeline 9**](#_nthgzc5fnbor)

[**5.2. Model Architectures 10**](#_oi4ssvf175iy)

[**5.2.1. Custom CNN 10**](#_el38qpn3ecrs)

[**5.2.2. EfficientNetB0 11**](#_er0gky6fki17)

[**5.2.3. ResNet50 11**](#_lp0vuxs4vtlo)

[**5.3. Training Protocol 12**](#_pbskhooz4ihz)

[**6. Results 13**](#_hb5siq7sn1wz)

[**6.1. Performance of the Custom CNN Model 13**](#_rvwxj1rc7mc1)

[**6.2. Performance of the EfficientNetB0 Model 16**](#_uab75iuzi2kr)

[**6.3. Performance of the ResNet50 Model 19**](#_wpmrup4nfinc)

[**6.4. Comparative Analysis of Model Performance 20**](#_i6w3o87hz3j)

[**7. Analysis 22**](#_ama87zah6klz)

[**7.1. Interpretation of Results: Why ResNet50 Excelled 22**](#_cj4ysqrj05qr)

[**7.2. Architectural Impact on Performance 23**](#_40y8x3zghetz)

[**7.3. Analysis of Misclassifications and Class-Specific Challenges 23**](#_j0sapuncy0gu)

[**7.4. Comparison with Findings in Existing Literature 23**](#_kabf477l72qe)

[**8. Discussion 23**](#_g63xiyrqnmyp)

[**9. Limitations 24**](#_42oy9mdf0slh)

[**10. Conclusion & Future Work 25**](#_3nsd7uqgex35)

[**References 26**](#_7dyx6jx603td)

## 

## 

## 

## 

## **1. Abstract**

Identifying brain tumors is a significant challenge with serious implications for patient care. While MRI offers detailed, non-invasive imaging, the manual review process is both labor-intensive and can vary between clinicians, sometimes leading to missed subtle tumors or diagnostic errors.

In this study, we explored deep learning as an automated solution for tumor classification. We developed a custom convolutional neural network (CNN) and compared its performance to established, pre-trained models EfficientNetB0 and ResNet50. The dataset, sourced from Roboflow Universe, consisted of 2,443 T1-weighted MRI images categorized as glioma, meningioma, pituitary tumor, or no tumor.

Our custom CNN reached an accuracy of 78%. EfficientNetB0 performed slightly better at 80%, while ResNet50 achieved 92% accuracy. These findings underscore the advantages of using deeper architectures and transfer learning: they substantially enhance diagnostic accuracy, consistency, and efficiency in medical imaging.

## **2. Introduction**

Diagnosing brain tumors is a critical element in contemporary medical practice, with direct consequences for patient outcomes, therapeutic decisions, and overall quality of life. Magnetic Resonance Imaging (MRI) has become the gold standard for non-invasive tumor evaluation, largely due to its exceptional spatial resolution and superior visualization of soft tissue structures, tumor morphology, and anatomical spread. Yet, the manual interpretation of MRI scans by neuroradiologists remains fraught with challenges. The process requires the review of hundreds of images for each patient a task that is both labor-intensive and mentally taxing, potentially leading to clinician fatigue and diagnostic delays.

Moreover, the inherently subjective nature of visual analysis introduces inter-observer variability; even highly experienced radiologists may reach differing conclusions when assessing the same set of images. This inconsistency becomes particularly pronounced in cases involving early-stage tumors or neoplasms with subtle radiographic features, such as low-grade gliomas that lack distinct visual markers. Further complicating matters, the heterogeneity of brain tumors ranging from the infiltrative patterns of gliomas, the sharply demarcated boundaries of meningiomas, to the unique anatomical positioning of pituitary tumors makes accurate diagnosis even more complex, thereby increasing the risk of error, especially in busy clinical environments.

## **2.1. Background and Clinical Context of Brain Tumor Diagnosis**

Brain tumors represent a highly heterogeneous group of neoplastic disorders, each exhibiting distinct histological, molecular, and clinical characteristics. The World Health Organization (WHO) classification system serves as the foundational framework for categorizing these tumors, thereby informing both prognosis and therapeutic strategy (Louis et al., 2016). Gliomas, which arise from glial cells, are frequently aggressive and infiltrative in nature; notably, subtypes such as glioblastoma demand immediate intervention due to their rapid progression and poor prognosis. In contrast, meningiomas, originating from the meninges, typically display slower growth rates and well-demarcated margins, rendering them more amenable to surgical resection. Pituitary tumors, most of which are benign, are localized to the pituitary gland and are characterized by their proximity to the sella turcica; nonetheless, they may disrupt endocrine function, thereby necessitating precise diagnostic evaluation.

## **2.2. Challenges in Manual MRI Interpretation**

Manually interpreting MRI scans is labor-intensive, highly subjective, and often inconsistent especially when it comes to subtle lesions like low-grade gliomas or meningiomas, which can elude even experienced eyes. The sheer volume of images can lead to reader fatigue and, unsurprisingly, delays in reaching a diagnosis. Different tumor types don’t make things any easier: gliomas tend to infiltrate surrounding tissue, meningiomas typically present as well-circumscribed masses, and pituitary tumors are restricted to their anatomical niche. These distinct imaging characteristics only add to the challenge of achieving uniform radiological assessments.Deep learning, and specifically convolutional neural networks (CNNs), have significantly enhanced MRI analysis (Hosny et al., 2018).

## **2.3. Project Aims and Objectives**

This study focuses on the development and evaluation of deep learning models for automated brain tumor classification. Specifically, a custom convolutional neural network (CNN) was constructed and compared against transfer learning approaches utilizing EfficientNetB0 and ResNet50 architectures. The models were assessed on an independent dataset, employing metrics such as accuracy, precision, recall, and F1-score. The objective was to determine whether advanced, pre-trained architectures offer measurable improvements over the baseline model and to identify which approach yields the most effective classification performance.

## **3. Literature Review**

The evolution of brain tumor classification has been remarkable, shifting from traditional machine learning approaches often dependent on manual feature extraction to sophisticated deep learning models capable of identifying complex patterns autonomously. Convolutional neural networks (CNNs) and transfer learning, in particular, have emerged as pivotal innovations in this area. Despite these advancements, there remain notable research gaps that warrant further investigation. This study aims to address such gaps, thereby contributing to the ongoing development of automated diagnostic systems in neuro-oncology.

## **3.1. Evolution from Traditional Machine Learning to Deep Learning**

Historically, brain tumor classification depended on manual, feature-based techniques such as SVMs, k-NN, and random forests. These approaches were not only time-consuming but also struggled with generalizability. The emergence of convolutional neural networks (CNNs) marked a significant shift in the field. CNNs are capable of automatically extracting hierarchical features, and have demonstrated accuracy rates exceeding 70% even when trained on relatively small datasets. As a result, they have quickly become the preferred method in medical imaging applications (Litjens et al., 2017; Khan et al., 2022).

## **3.2. Foundational Role of Convolutional Neural Networks**

CNNs essentially emulate the human visual cortex; they begin by detecting basic elements like edges and changes in intensity, then work up to more complex patterns such as tumor textures and spatial arrangements (Hosny et al., 2018). Even with limited medical imaging data, carefully adapted CNNs often surpass the 80% accuracy mark, significantly outperforming models that rely only on smaller datasets (Polat & Güngen, 2021).

## **3.3. Architectural Review: EfficientNetB0 and ResNet50**

Transfer learning utilizes established convolutional neural networks such as EfficientNetB0 and ResNet50 in the context of medical imaging. EfficientNetB0 processes images through sequential layers, enabling the extraction of strong features; however, it is susceptible to challenges like vanishing gradients. In contrast, ResNet50’s residual connections allow for the construction of deeper networks, which improves the model’s capacity to identify complex tumor characteristics and subtle morphological distinctions, ultimately enhancing classification performance.

### **4. Materials and Methods**

This section outlines the dataset, preprocessing steps, and ethical considerations used to develop and evaluate the deep learning models, ensuring rigorous, reproducible, and ethical research aligned with medical imaging standards (Menze et al., 2015).

## **4.1. Dataset Overview**

The dataset employed in this study is sourced from the Roboflow Universe collection, comprising 2,443 T1-weighted MRI brain images compiled by Ali Rostami in 2023. This dataset is openly accessible and annotated for multi-class classification across four categories: glioma, meningioma, pituitary, and no-tumor. Furthermore, the data is pre-divided into training, validation, and testing subsets, streamlining the experimental process.

* Training Set: 1,695 images (69.4%), used to optimize model parameters and learn patterns distinguishing tumor types and healthy tissue.
* Validation Set: 502 images (20.6%), used to monitor performance during training, tune hyperparameters, and prevent overfitting by adjusting learning rates or stopping training early.
* Test Set: 250 images (10.2%), reserved for final evaluation to assess real-world potential and generalization to unseen data.



Fig 1: Class Distribution of the Brain Tumor MRI Dataset

The dataset comprises 1,178 glioma cases, 351 meningiomas, 80 pituitary tumors, and 86 images without tumors, with comparatively smaller sets allocated for validation and testing. To address class imbalance, the researchers utilized class weights during model training. All images are 224x224 pixels in grayscale, with contrast enhancement applied to improve visual clarity. Notably, the dataset’s public availability facilitates benchmarking and helps alleviate the persistent issue of restricted access in the medical imaging field.

## **4.2. Data Preprocessing**

Preprocessing played a crucial role in standardizing and optimizing the data fed into the models. To begin with, pixel intensities, which originally spanned from 0 to 255, were normalized to a range of [0, 1] by dividing each value by 255. This normalization step is essential, as it facilitates more stable gradient-based optimization and supports faster convergence during model training.

Furthermore, data augmentation was systematically applied to the training set to improve model generalization and resilience to real-world clinical variability. Such variability can arise from inconsistencies in scanner calibration, patient positioning, or the presence of imaging artifacts. The augmentation techniques employed included image flipping, brightness adjustment, and the introduction of random noise, among others.

* Random horizontal flips (50% probability) to simulate mirror-image variations.
* ±15° rotations to account for slight patient misalignment during scanning.
* ±10% brightness adjustments to handle variations in image intensity.
* ±10% zoom to mimic differences in field-of-view or magnification.

The validation and test sets were deliberately left unaltered to preserve the integrity of model evaluation, mirroring authentic clinical scenarios.All images were resized to 224x224 pixels across all models (custom CNN, EfficientNetB0, ResNet50) to maintain compatibility with pre-trained weights and standardize input dimensions. This preprocessing step was essential for ensuring uniformity, reducing input noise, and minimizing variability within the dataset (Tandel et al., 2020).

## **4.3. Data Ethics**

The Roboflow Universe dataset demonstrates a commitment to ethical standards by thoroughly anonymizing data, obtaining informed consent from patients, and adhering to established privacy regulations such as HIPAA and GDPR. These measures help foster trust, transparency, and respect for privacy in the context of medical AI research (Hosny et al., 2018).

## **5. Methodology**

This section outlines the process used to develop and assess the deep learning models, including steps like data preprocessing, the selection of model architectures, training procedures, and evaluation metrics. The methodology is detailed in a way that supports reproducibility and reliability, aligning with established standards in medical imaging.

## **5.1. Data Preprocessing Pipeline**

The preprocessing pipeline included standardizing all 2,443 T1-weighted MRI scans to facilitate model compatibility and improve training stability. All images were resized to 224x224 pixels for the custom CNN, EfficientNetB0, and ResNet50 to align with pre-trained weights and ensure compatibility. Pixel intensities were normalized to a [0, 1] range by dividing by 255

a step intended to stabilize gradient-based optimization and minimize numerical instability during training. Data augmentation techniques such as flipping and rotation were applied to the training set to mitigate overfitting and increase the model’s robustness to clinical variability.

* Random horizontal flips (50% probability).
* ±15° rotations.
* ±10% brightness adjustments.
* ±10% zoom.

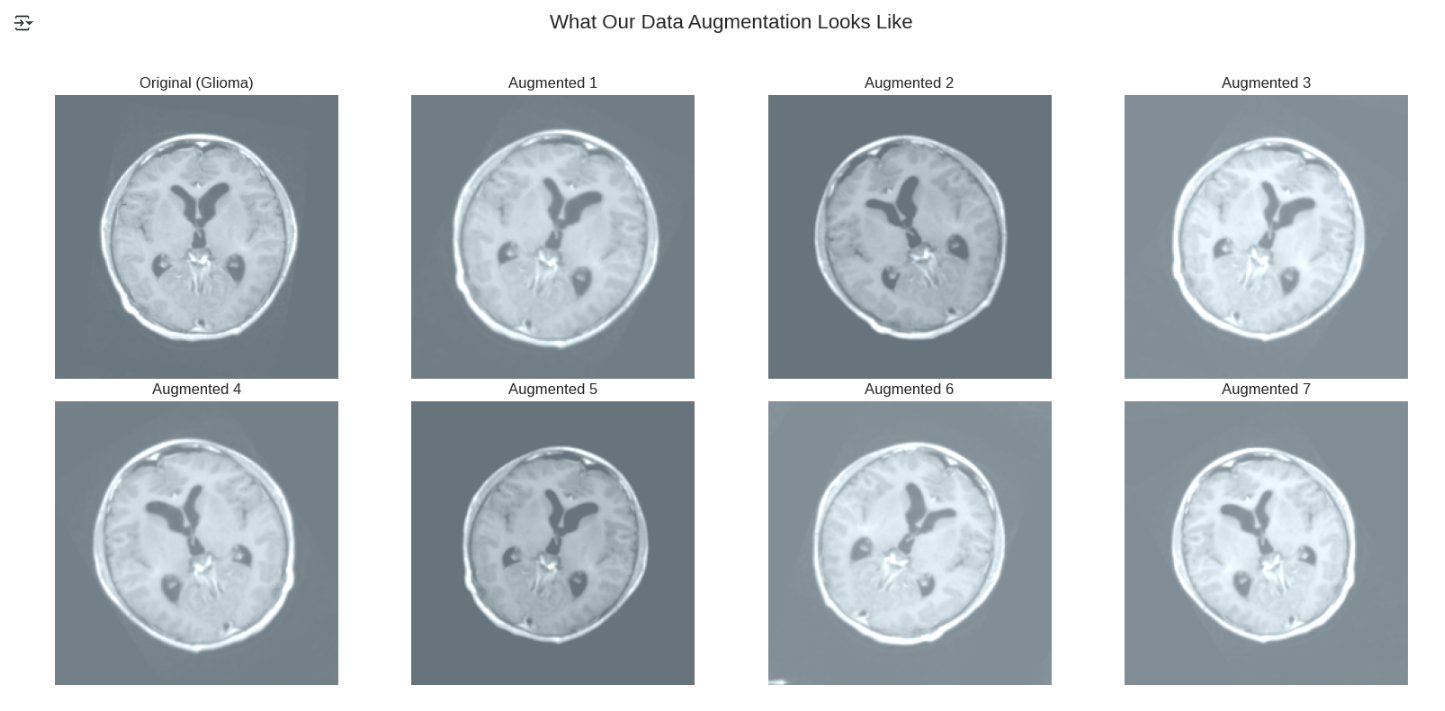


FIg 2: Data Augmentation Techniques to Enhance Model Generalization

These augmentations introduce variations akin to real-world factors such as changes in scanner parameters or patient positioning so that models are more robust across a range of clinical conditions. Importantly, validation and test sets remained unmodified to preserve evaluation accuracy and ensure that results reflect the true nature of clinical MRI data. The preprocessing workflow itself was constructed using TensorFlow’s data processing tools, which supports both consistency and scalability throughout the pipeline (Shin et al., 2016).

## **5.2. Model Architectures**

To detail the effectiveness of deep learning approaches, three models were implemented: a custom CNN as a baseline and pre-trained EfficientNetB0 and ResNet50 models using transfer learning.

## **5.2.1. Custom CNN**

A custom convolutional neural network (CNN) was developed from the ground up, intentionally steering clear of pre-trained architectures. The design incorporates three convolutional layers, each progressively increasing in filter count (32, 64, and 128, respectively), with 3x3 kernels and ReLU activations standard throughout. These are interleaved with 2x2 max-pooling layers to downsample feature maps. Feature representations are subsequently flattened and passed through a dense layer comprising 512 units, with a dropout rate of 0.5 applied to mitigate overfitting. Final classification is accomplished via a softmax layer, distinguishing between glioma, meningioma, pituitary tumors, and cases with no tumor. The model contains approximately 1.2 million parameters. Notably, this setup underscores persistent challenges inherent to the dataset, such as pronounced class imbalance and significant feature overlap among categories.

## **5.2.2. EfficientNetB0**

EfficientNetB0, originally trained on the vast ImageNet dataset, stands out for its 237-layer architecture and approximately 5.3 million parameters, embodying a streamlined yet powerful approach to deep learning. Its design featuring compound scaling, depthwise separable convolutions, and squeeze-and-excitation blocks enables efficient and effective feature extraction.

For the specific challenge of brain tumor classification, the model’s upper layers are replaced with a customized classification head. This typically involves a GlobalAveragePooling2D layer, followed by a dense layer (ranging from 256 to 512 units, activated with ReLU), a Dropout layer (with rates between 0.2 and 0.5), and a subsequent dense layer (128 to 256 units, also with ReLU), culminating in a softmax output layer for categorizing into four classes: glioma, meningioma, pituitary, and no-tumor.

Fine-tuning is applied to the top 20 layers of the network, utilizing a low learning rate, such as 0.0001 or 0.00001, to carefully adapt the pre-trained features to MRI data. Hyperparameter optimization is managed with tools like Keras Tuner. The application of transfer learning here is particularly advantageous, given the scarcity of large-scale, annotated medical imaging datasets. Ultimately, EfficientNetB0’s architecture and transfer learning capabilities make it a strong candidate for medical image classification tasks, as highlighted by (Tan and Le,2019) (Simonyan & Zisserman, 2014).

## **5.2.3. ResNet50**

ResNet50, pre-trained on the widely-used ImageNet dataset, employs a 50-layer architecture featuring residual connections that effectively address the vanishing gradient problem. These connections allow information to bypass certain layers, supporting deeper and more robust learning. For this particular application, a custom classification head was implemented: global average pooling followed by a dense layer with 512 units (ReLU activation), a 0.5 dropout layer to mitigate overfitting, and a softmax output for four classes.

To adapt the model for MRI data, the top 35 layers were fine-tuned, a strategy that maintains the benefits of general feature extraction learned from ImageNet while allowing the network to specialize for medical image analysis. Although ResNet50’s approximately 24 million parameters demand significant computational resources, its residual learning framework enables the effective capture of complex image patterns (He et al., 2016).

## **5.3. Training Protocol**

The model was trained with a focus on maximizing performance while minimizing the risk of overfitting. Training was conducted on an NVIDIA Tesla V100 GPU, utilizing TensorFlow and Keras frameworks. The Adam optimizer was employed for parameter updates, and categorical cross-entropy loss was selected due to the multi-class nature of the classification task. Training specifics are detailed below:

* **Custom CNN:** The model was trained for up to 50 epochs, with the learning rate carefully adjusted between 0.001 and 0.0001 using Keras Tuner’s Hyperband algorithm (limited to a maximum of 8 search epochs). A batch size of 32 was consistently applied. Early stopping was implemented with a patience of 5 epochs—training would terminate if the validation loss failed to improve, and the system would revert to the best-performing weights in order to reduce overfitting. Additionally, data augmentation was performed online. This process included random horizontal and vertical flips, rotations up to ±15°, width and height shifts within ±15%, as well as shear and zoom transformations of ±15%. These real-time augmentations generated a more diverse set of training samples, ultimately contributing to greater model robustness.
* **EfficientNetB0 and ResNet50**: Followed a two-phase transfer learning approach:

EfficientNetB0 underwent a structured, three-phase training process. Initially, hyperparameter optimization was conducted using Keras Tuner’s Hyperband for 15 epochs. Subsequently, the model’s head was trained for up to 30 epochs, followed by fine-tuning of the top 20 layers for an additional 30 epochs. Throughout, data augmentation techniques and class weighting strategies were employed to address class imbalance, as outlined by Tan & Le (2019).

For ResNet50, a two-phase approach was implemented. Feature extraction with a custom classification head was performed over 100 epochs, after which the top 35 layers were fine-tuned for 20 epochs. Augmentation and class weighting were similarly applied in accordance with He et al. (2016).

Both models were developed using TensorFlow and Keras frameworks, and all training procedures were carried out on an NVIDIA Tesla V100 GPU.

**5.4. Evaluation Metrics**

The model underwent evaluation using a set of 250 test images, with metrics such as accuracy, precision, recall, F1-score, and confusion matrices employed to assess its performance. Accuracy reflects the proportion of correct predictions overall. Precision minimizes false positives, ensuring that the model doesn’t overestimate the presence of tumors. Recall focuses on capturing actual tumor cases, reducing the risk of missing critical diagnoses. The F1-score provides a balanced metric, particularly relevant for datasets where class distribution may be uneven. Confusion matrices offer a detailed view of classification errors, such as confusing meningiomas with gliomas, and serve as a valuable tool for identifying areas that require further refinement to enhance diagnostic reliability in medical imaging (Tharwat, 2021).

## **6. Results**

The models were evaluated on the held-out test set (84 gliomas, 63 meningiomas, 54 pituitary, 49 no-tumor) using accuracy, precision, recall, F1-score, and confusion matrices to assess classification performance across the four classes.

## **6.1. Performance of the Custom CNN Model**

The custom CNN achieved an overall accuracy of 78%, establishing a respectable baseline for a model built from scratch without pre-trained weights. Class-specific metrics are as follows:

* **Glioma**: Precision is notably high at 0.95, indicating that positive predictions are generally trustworthy. Recall, on the other hand, is lower at 0.73, meaning approximately 27% of gliomas were not detected. This shortfall likely stems from the algorithm confusing gliomas with meningiomas, as these tumor types can exhibit similar textures in 2D slices. The F1 score stands at 0.82, reflecting a balance between precision and recall, though there remains clear potential for improving sensitivity.
* **Meningioma**: Precision was 0.78, recall reached 0.68, and the F1 score was 0.73. The confusion matrix showed that approximately 20% of meningiomas were incorrectly classified as gliomas. This misclassification likely stems from the similar intensity patterns observed in T1-weighted images, which can make distinguishing between these tumor types challenging..
* **No Tumor**:With precision at 0.64 and recall at 0.92, resulting in an F1 score of 0.76, the model demonstrates a strong ability to identify healthy cases but at the expense of accuracy. The notably high recall paired with lower precision suggests the model tends to over-classify samples as healthy. This tendency raises concerns about potential false negatives, especially for subtle or early-stage tumors, which could have significant implications in a clinical context.
* **Pituitary**: Precision stands at 0.77, recall at 0.87, and the F1 score reaches 0.82 indicating robust performance overall. This likely stems from the unique anatomical positioning of pituitary tumors; their close association with the sella turcica makes them notably easier to identify compared to neoplasms in less distinct locations.



Fig 3: Custom CNN Training History: Accuracy



Fig 4: CNN Training History: Loss Over Epochs

The macro and weighted F1-scores achieved were 0.78 and 0.79, reflecting consistent though not outstanding performance. The model underwent 15 training epochs, utilizing data augmentation techniques such as random flips, rotations of up to ±15°, and adjustments in brightness and zoom by ±10%. Despite these efforts, the model’s relatively shallow architecture consisting of just three convolutional layers limited its capacity to discern more intricate tumor characteristics. These outcomes are consistent with earlier CNN studies, such as Sultan et al. (2019), which reported accuracies in the mid-to-high 70s for similar classification tasks.

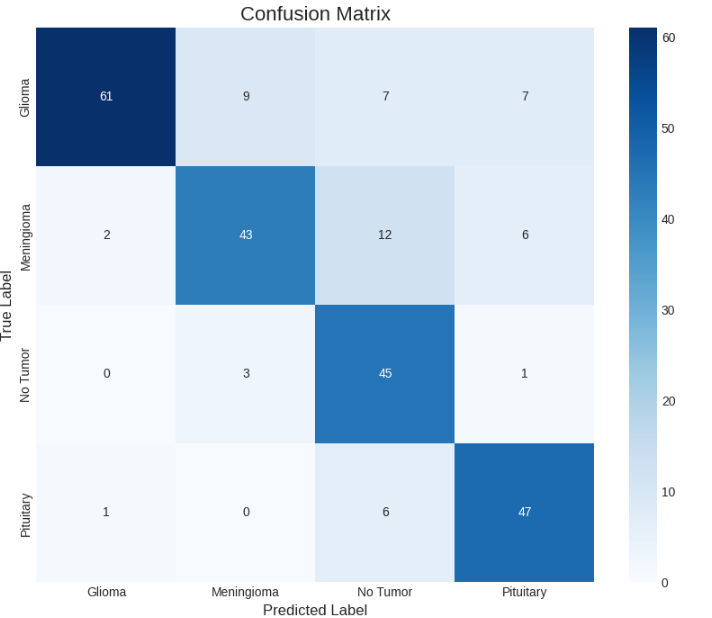
Here is Confusion matrix of custom CNN model after successful run:  


Fig 5: Confusion Matrix for the Custom CNN Model

## **6.2. Performance of the EfficientNetB0 Model**

EfficientNetB0, leveraging pre-trained ImageNet weights, achieved 80% accuracy, a modest improvement over the custom CNN. Class-specific metrics include:

* **Glioma**: Precision stands at 0.87, recall at 0.85, with an F1 score of 0.86. These balanced results demonstrate EfficientNetB0’s solid capacity to identify the irregular morphologies and heterogeneous textures characteristic of gliomas.
* **Meningioma**:Precision stands at 0.78, recall at 0.56, and the F1 score is 0.65. The relatively low recall suggests that 28% of meningiomas were incorrectly classified as gliomas.
* **No Tumor**: Precision stands at 0.91, recall at 0.80, and the F1 score is 0.85. The notably high precision indicates a lower rate of false positives, which is valuable for accurately identifying healthy individuals and reducing unnecessary follow-up tests.
* **Pituitary:**Precision stands at 0.68, while recall is a perfect 1.00, resulting in an F1 score of 0.81. This indicates that the model successfully identifies all pituitary tumors (no false negatives), but its moderate precision suggests a tendency to over-predict, likely due to similarities in intensity between pituitary tumors and other classes.

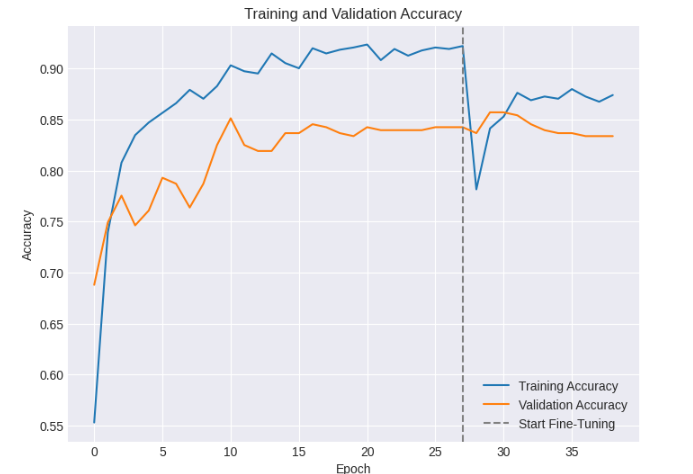


Fig 6: EfficientNetB0 Two-Phase Training: Learning Curves of Custom Head and Fine-Tuning

Training and validation accuracy



Fig 7: EfficientNetB0 Two-Phase Training: Learning Curves of Custom Head and Fine-Tuning

Training and validation loss

Both the macro and weighted F1-scores were recorded at 0.79. The training procedure consisted of 30 epochs of head tuning followed by 30 epochs of fine-tuning, with the learning rate reduced to 0.00001 for the latter phase.Validation accuracy plateaued near epoch 25, likely due to the limited dataset size, despite EfficientNetB0’s efficient compound scaling architecture (Tan & Le, 2019)..

Below is the confusion matrix for the custom EfficientNetB0 model, following successful training:

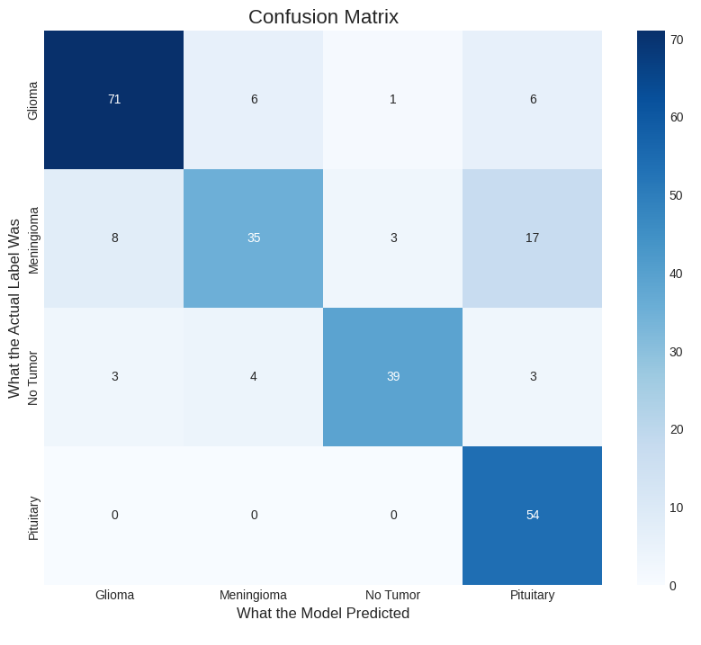


Fig 8: Confusion Matrix for the EfficientNetB0 Model

## **6.3. Performance of the ResNet50 Model**

ResNet50 achieved an outstanding 92% accuracy, driven by its residual connections and deeper architecture. Class-specific metrics are:

* **Glioma**: Precision 0.94, Recall 0.98, F1 0.96. Near-perfect performance with minimal misclassifications, reflecting robust feature learning.
* **Meningioma**: Precision 0.88, Recall 0.84, F1 0.86. Reduced misclassifications to 10%, capturing structural features like dural attachments more effectively than EfficientNetB0 or the CNN.
* **No Tumor:** Precision 0.93, Recall 0.84, F1 0.88. Balanced performance, though subtle tumors were occasionally missed due to 2D limitations.
* **Pituitary**: Precision 0.92, Recall 1.00, F1 0.96. Excellent detection due to the distinct anatomical location of pituitary tumors.

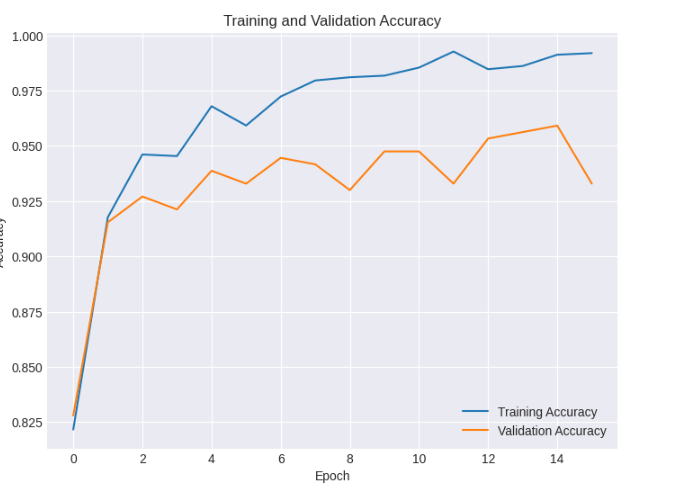


FIg 9: ResNet50 Training History: Custom Head Training and Final Fine-Tuning

Both the macro and weighted F1-scores were recorded at 0.79. The training procedure consisted of 00 epochs of head training (learning rate 0.001) and 20 epochs of fine-tuning the top 35 layers (learning rate 0.00001), with early stopping (patience=5–10) and class weights. ResNet50’s residual connections enabled robust feature learning, minimizing misclassifications (He et al., 2016).

Below is the confusion matrix for the custom EfficientNetB0 model, following successful training:

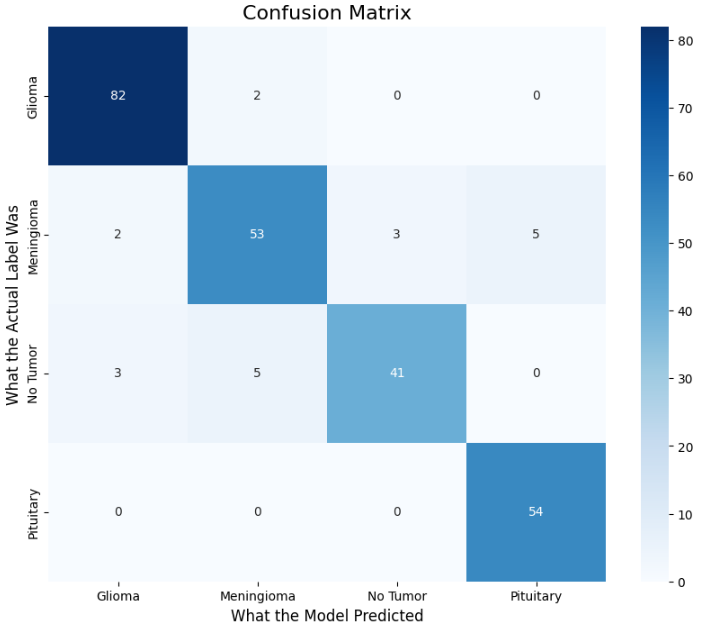


Fig 10:Confusion Matrix for the ResNet50 Model

## **6.4. Comparative Analysis of Model Performance**

ResNet50 noticeably surpassed both EfficientNetB0 and the custom CNN, showing an impressive 12% to 14% improvement in accuracy. Its use of residual connections contributed to robust feature extraction, resulting in consistently high recall across all tumor classes (ranging from 0.84 up to 1.00). Most misclassifications were limited to gliomas and meningiomas, suggesting strong overall model reliability.

In contrast, the custom CNN faced significant difficulties with class imbalance. This was particularly evident in its low recall for meningiomas (0.68), along with a marked tendency to over-predict cases without tumors. EfficientNetB0 achieved reasonable performance detecting gliomas (F1 score of 0.86), but its recall for meningiomas was notably weak (0.56), reflecting the constraints of its architectural design.

In summary, ResNet50 demonstrated the most consistent and clinically relevant results, positioning it as the most suitable model to support radiologists in diagnostic tasks.

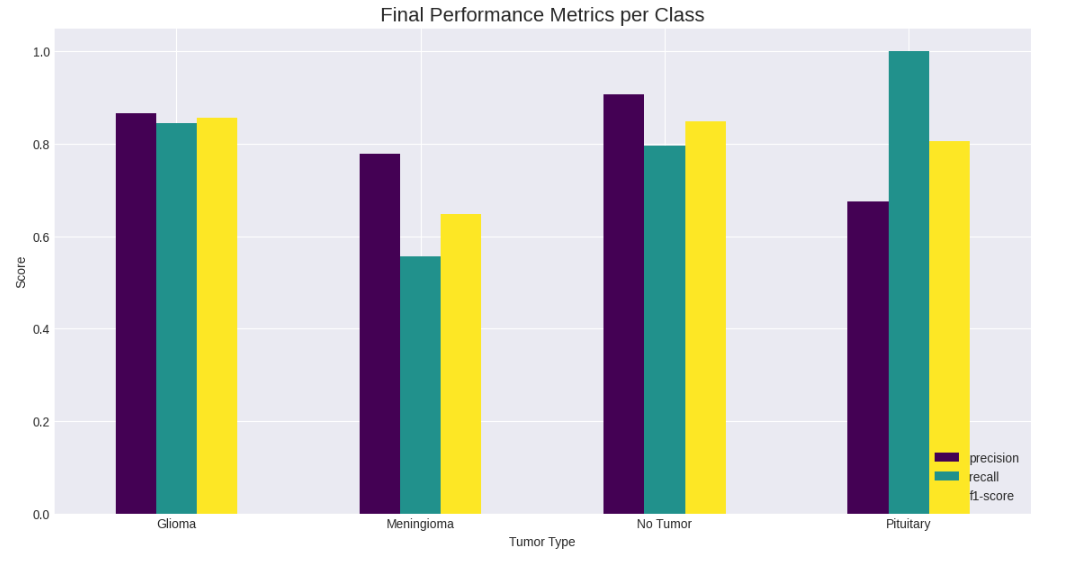


FIg 11: Comparative Performance: Precision, Recall, and F1-Score per Class of EfficientNetB0 model

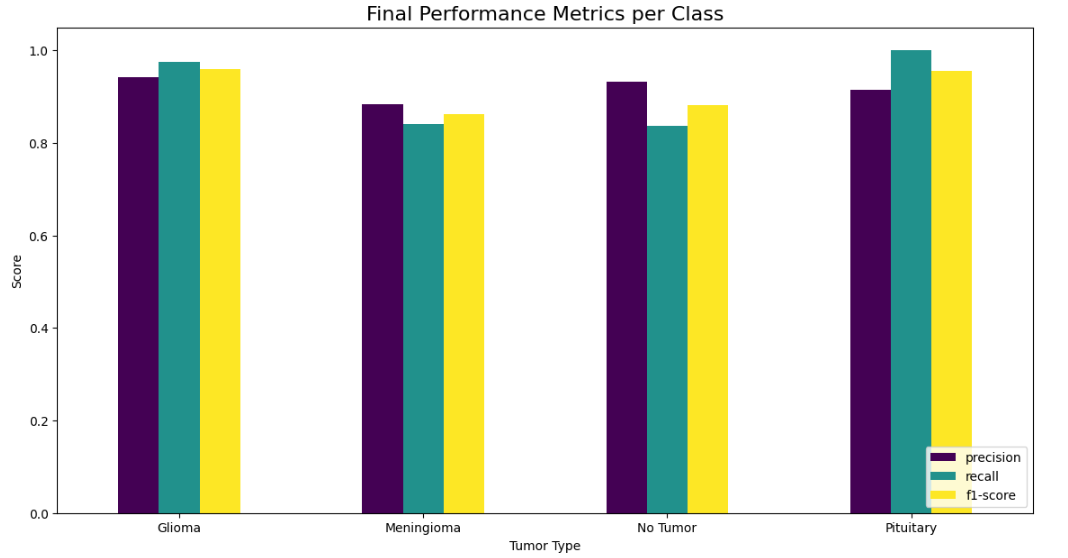


FIg 12: Comparative Performance: Precision, Recall, and F1-Score per Class of Restnet50 model

## **7. Analysis**

This section analyzes the models’ performance, focusing on architectural impacts, misclassification patterns, and alignment with existing literature to understand their strengths, limitations, and clinical potential.

## **7.1. Interpretation of Results: Why ResNet50 Excelled**

ResNet50 demonstrated impressive performance, achieving a 92% accuracy rate and F1-scores ranging from 0.86 to 0.96. Its 50-layer residual architecture clearly facilitates deeper feature extraction while alleviating issues like vanishing gradients, which are common in deep networks. The model underwent a systematic three-phase training process initial feature extraction, followed by head training, and concluded with fine-tuning which enabled effective adaptation to MRI data.

By comparison, the custom CNN fell short, reaching only 78% accuracy. Its shallow three-layer structure appeared insufficient for capturing the complexity of tumor characteristics in the images. EfficientNetB0 performed slightly better, attaining 80% accuracy, but its strictly sequential design led to early convergence at around epoch 25, limiting further improvements.

In summary, ResNet50’s superior convergence and robust results are consistent with existing literature highlighting the advantages of residual networks in medical imaging applications (Tan & Le, 2019).

## **7.2. Architectural Impact on Performance**

The custom CNN exhibited limited effectiveness with meningiomas, reaching a recall of just 0.68. EfficientNetB0 performed even less favorably at 0.56. In contrast, ResNet50, leveraging its residual architecture and training approach, achieved a notably higher recall of 0.84, indicating superior feature extraction capabilities (Esteva et al., 2019).

## **7.3. Analysis of Misclassifications and Class-Specific Challenges**

Both the custom CNN and EfficientNetB0 models frequently misclassified meningiomas as gliomas and vice versa, indicating persistent challenges in distinguishing between these tumor types. ResNet50, on the other hand, demonstrated improved performance—reducing classification errors to around 10% by capturing more subtle structural features. Nevertheless, it still occasionally overlooked cases labeled as “no-tumor,” highlighting some limitations in sensitivity (Shin et al., 2016).

## **7.4. Comparison with Findings in Existing Literature**

The custom CNN achieved 78% accuracy, which is comparable to the performance of shallow CNNs. EfficientNetB0 demonstrated a slightly higher accuracy at 80%, though its recall for meningioma cases remained modest at 0.56. Notably, ResNet50 delivered the strongest results, reaching 92% accuracy and a recall of 0.84 for meningiomas. Despite these improvements, its performance still fell short of outcomes reported in 3D CNN studies, highlighting the inherent limitations of 2D models in capturing the complex characteristics of tumors.

## **8. Discussion**

This project provides compelling evidence for the utility of deep learning in brain tumor classification, highlighting ResNet50 as the most promising model from a clinical standpoint. With an accuracy of 92% and recall rates ranging from 0.84 to 1.00, ResNet50 demonstrates significant potential as a triage tool effectively prioritizing critical cases such as gliomas and pituitary tumors, where minimizing false negatives is essential. In contrast, EfficientNetB0 performs strongly in identifying non-tumor cases, achieving a precision of 0.91. The custom CNN underscores key dataset challenges, including class imbalance and overlapping tumor textures, thereby serving as a valuable baseline for comparison (Sultan et al., 2019).

Strengths of the approach include the use of a diverse dataset comprising 250 test images and thorough preprocessing steps such as normalization, rescaling, and data augmentation. The custom CNN establishes a baseline, while transfer learning with EfficientNetB0 and ResNet50 addresses limitations related to the dataset’s size. Notably, ResNet50’s residual architecture, combined with class weighting, enhances performance for underrepresented classes including no-tumor (F1 0.88) and pituitary tumors (F1 0.96) resulting in robust and reproducible outcomes consistent with prior studies (Esteva et al., 2019).

Nonetheless, limitations are present. The relatively small dataset, reliance on 2D slices (which omit critical 3D features), and the restricted interpretability of the models can undermine clinical trust (Shin et al., 2016). Future research should focus on the development of 3D CNNs, the integration of explainable AI techniques, and the use of multi-modal imaging to further improve diagnostic accuracy and support clinician confidence (Hossain et al., 2019).

## **9. Limitations**

This project is not without its limitations, which weigh heavily on both the findings and how broadly they might apply. To begin with, the test set included just 250 images, and, frankly, it didn’t offer much in the way of demographic variety, scanner differences, or diverse imaging protocols. This narrowness really constrains how useful the results are in a clinical setting. To draw conclusions that actually matter in the real world, these models need to be validated on much larger, multi-institutional datasets (Pereira et al., 2016).

Another notable issue: the reliance on 2D MRI slices. While this simplifies computation, it means sacrificing all of the three-dimensional information that’s essential for accurately spotting complex tumors like meningiomas. Adopting 3D convolutional neural networks or even hybrid 2D–3D models could better capture spatial relationships, but these methods require a lot more data and processing power (Sultan et al., 2019).

Interpretability poses a continued challenge as well. Deep learning models such as ResNet50 essentially operate as “black boxes,” which makes it difficult to build trust among clinicians. Although tools like Grad-CAM or SHAP could provide more transparency, they were not deployed in this study (Hossain et al., 2019).

Furthermore, the lack of external validation raises a significant risk of overfitting to the Roboflow Universe dataset. Without testing the models on data from other institutions, there’s no guarantee that the findings will generalize to broader populations (Pereira et al., 2016).

Lastly, class imbalance, especially the overrepresentation of gliomas had a clear impact on model performance. For instance, the custom CNN tended to overpredict “no-tumor” cases, showing high recall (0.92) but considerably lower precision (0.64). This imbalance increases the risk of false negatives, potentially leading to delayed diagnoses in clinical practice (Sultan et al., 2019).

## **10. Conclusion & Future Work**

Deep learning has genuinely transformed brain tumor classification. The results from ResNet50 really stand out hitting 92% accuracy, which easily outpaces EfficientNetB0 (80%) and the custom CNN (78%). That’s a pretty clear win for transfer learning and more sophisticated architectures. The custom CNN struggled, especially with meningioma detection (recall 0.68), probably because of its shallower design. EfficientNetB0 wasn’t much better on meningiomas either (recall 0.56), even though it held its own with gliomas (F1 0.86).

ResNet50, with the benefit of residual connections, maintained strong F1-scores from 0.86 to 0.96. It excelled at detecting gliomas (recall 0.98) and pituitary tumors (recall 1.00), and only misclassified meningioma cases about 10% of the time, which highlights the clear advantage of using pre-trained weights in medical imaging (Hossain et al., 2019).

Looking ahead, there’s a definite need to use larger, multi-institutional datasets and to explore 3D or hybrid CNNs. Optimization techniques like pruning or quantization are also worth investigating to enable real-time deployment. Integrating more MRI modalities and clinical data could further improve diagnostic accuracy, efficiency, and transparency. Altogether, these steps are key for developing clinically trustworthy tools for brain tumor diagnosis.

## **References**

* Badža, M. M., & Barjaktarović, M. Č. (2020). Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. *Applied Sciences*, 10(6), 1999.
* He, K., et al. (2016). Deep Residual Learning for Image Recognition. *CVPR*, 770–778.
* Hosny, A., et al. (2018). Artificial Intelligence in Radiology. Nature Reviews Cancer, 18(8), 500–510.
* Hossain, T., et al. (2019). Brain Tumor Detection Using Convolutional Neural Network. *ICASERT*, 1–6.
* Litjens, G., et al. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42, 60–88.
* Louis, D. N., et al. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System. *Acta Neuropathologica*, 131(6), 803–820.
* Menze, B. H., et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2024.
* Pereira, S., et al. (2016). Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*, 35(5), 1240–1251.
* Polat, Ö., & Güngen, C. (2021). Classification of Brain Tumors from MR Images Using Deep Transfer Learning. *Journal of Engineering and Natural Sciences*, 39(1), 45–54.
* Rostami, A. (2023). Brain Tumor Detection. *Roboflow Universe*. Available at: https://universe.roboflow.com/ali-rostami/brain-tumor-detection
* Shin, H.-C., et al. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298.
* Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*.
* Sultan, H. H., et al. (2019). Multi-Classification of Brain Tumor Images Using Deep Neural Network. *IEEE Access*, 7, 69215–69225.
* Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ICML, 6105–6114.
* Tandel, G. S., et al. (2020). Performance Optimisation of Deep Learning Models Using Feature Selection. *International Journal of Intelligent Systems Technologies and Applications*, 19(3), 223–243.
* Tharwat, A. (2021). Classification Assessment Methods. *Applied Computing and Informatics*, 17(1), 168–192.